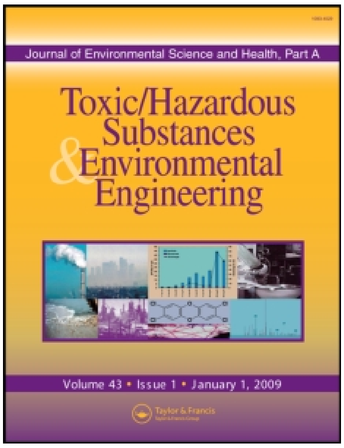


Journal of Environmental Science and Health, Part A
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597268

Geostatistical simulation and estimation of the spatial variability of soil zinc

Yu-Pin Lin [a]; Tsun-Kuo Chang [b]
[a] Department of Landscape Architecture, Chinese Culture University, Taipei, Taiwan, R.O.C. [b] Graduate Institute of Agricultural Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

To cite this Article Lin, Yu-Pin and Chang, Tsun-Kuo'Geostatistical simulation and estimation of the spatial variability of soil zinc', Journal of Environmental Science and Health, Part A, 35: 3, 327 — 347
To link to this Article: DOI: 10.1080/10934520009376974
URL: http://dx.doi.org/10.1080/10934520009376974

J. ENVIRON. SCI. HEALTH, A35(3), 327–347 (2000)

GEOSTATISTICAL SIMULATION AND ESTIMATION OF THE SPATIAL VARIABILITY OF SOIL ZINC

Key Words: Geostatistical simulation, kriging, soil zinc, spatial variability, Geographic Information Systems

Yu-Pin Lin[1]* and Tsun-Kuo Chang[2]

[1]Department of Landscape Architecture, Chinese Culture University
Taipei, Taiwan 11114 R.O.C.
[2]Graduate Institute of Agricultural Engineering, National Taiwan University
Taipei, Taiwan 10617 R.O.C.

ABSTRACT

Collected data in soil heavy metal investigations may contain significant levels of uncertainty, including complex and even unexplainable spatial variations at a small investigation site. Therefore, this study identifies the spatial structure of soil zinc in the northern part of Changhua County in Taiwan to understand the spatial variation and uncertainty of soil zinc. The spatial maps of this heavy metal are simulated by using the geostatistical simulation, and estimated by using ordinary kriging and natural log kriging. The estimation and simulation results indicate that Sequential Gaussian Simulations can reproduce the spatial structure for investigated data. Furthermore, displaying a low spatial variability, the ordinary kriging and natural log kriging estimates can not fit the spatial structure and small-scale variation for the soil zinc investigated data. The maps of kriging estimates are much smoother than those of simulations. Sequential Gaussian

* Corresponding author; e-mail: yplin@ccu016.pccu.edu.tw

Simulation with multiple realizations has significant advantages at a site with high variation investigated data over ordinary kriging, even natural log kriging techniques. Geographic information systems display these simulation and estimation results.

INTRODUCTION

Thoroughly understanding the spatial characteristics of soil pollutants ensures the effectiveness of risk assessment and remediation. These characteristics of soil investigated data occasionally display a complicated spatial variation over investigated site. Therefore, the spatial distribution of these soil pollutants must be characterized. Geostatistics may be defined as a collection of techniques for solving estimation problems related to spatial variables (Journel and Huijbregts, 1978, and ASCE 1990). A geostatistcal technique, kriging is a linear interpolation procedure. It provides a best linear unbiased estimator (BLUE) for quantities that vary in space. Geostatistical approaches have recently been applied to analyze the spatial variability of pollutants, with notable examples including Yost et al. (1982), Keck et al. (1993), Samra and Gill (1993), Litaor (1995), Couto et al. (1997), Juang and Lee (1998), Wang (1998), Wang and Zhang (1999), and White et al. (1999).

The kriging process yields weighted-average estimates that may fail to preserve the variability of the investigated process. Minimizing the prediction error variance involves smoothing the actual variability (Journel and Huijbergts, 1978). The estimated values based on kriging display a lower variation than the actual investigated values. To correct the above deficiency, geostatistical simulation can be performed. Simulation generates equally likely sets of values for a variable, which are consistent with available in-situ measurements. This often implies that the simulated values have the same mean and the same variogram as the original data; they may also have to coincide with the original data at measurement points. Simulation focuses mainly on reproducing the fluctuations in the observations, instead of producing the optimal prediction (Sterk and Stein, 1997).

On the other hand, soil investigated data occasionally possess a skewed distribution even lognormal distribution. Therefore, lognormal kriging was developed in geostatistics to account for the frequently observed skewed distribution of the investigated data (Roth, 1998). This technique transforms the data into lognormal formation before kriging estimation. Notable works include Rendu (1979), Journel (1980), Dowd (1982), Rivoirard (1990), and Roth (1998).

Conditional simulation attempts not only to generate a set of values that have some specified mean and covariance, but also to reproduce observed data at several locations. Conditional simulations are useful in many instances. For example, assume that a variable is investigated at several locations. Such measurements can be used, along with simulated values, to analyze the spatial distribution of the variable in question. Eggleston et al. (1996) used conditional simulation and ordinary kriging to reproduce hydraulic conductivity structure and sensitivity under limited amounts of data. Later, Mowrer (1997) used sequential gaussian simulation to create maps of potential old-growth forest conditions across a 121 hectare first-order subalpine watershed. More recently, Kentwell et al. (1999) used sequential gaussian fractal simulation to enhance the prediction accuracy of grade tonnage curve.

In this study, we employ ordinary kriging, natural log (ln) ordinary kriging and conditional simulation techniques to produce the maps and realizations of Zn in a case study. The descriptive statistics, spatial structure (experimental variogram), and spatial structure of estimated and simulated results are also discussed and compared in this study. Finally, the estimation and simulation results are also displayed in geographic information systems (GIS).

MATERIALS AND METHODS

Data

The Environmental Protection Administration (EPA) of the Republic of China initiated a collaborative research program in 1983 to determine the As, Cd, Cu, Cr, Hg, Ni, Pb and Zn trace elements of soil and other soil properties such as cation-exchange capacity, and pH. In those related studies, soils were sampled from 878 sites representing important agricultural production areas across Taiwan in reports on the elemental contents of soils in Taiwan.

This study develops maps illustrating the geographic distribution of Zn in surficial soil horizons in the northern part of Changhua County in Taiwan using geostatistical estimation, simulation and geographic information systems (GIS). Data were derived from the EPA studies described above. In this study area, 350 sampling points (Figure 1) are used for estimating and simulating the spatial distribution of Zn. These measurements were conducted from 1983 through 1986. Samples were taken from geographically distributed sites at a target interval of 4 km network. Soils were sampled at a depth of 0-15 cm. Moreover, Zn in the filtrates was determined by atomic absorption spectroscopy.

Table 1 summarizes the descriptive statistics of Zn investigated data. Mean Zn and ln(Zn) for complete data are 80.697(mg/kg) and 3.922(Ln(mg/kg)), and the variance of Zn and LnZn are 1246.405((mg/kg)^2) and 0.787((Ln(mg/kg)^2). The spatial maps of Zn and ln(Zn) are displayed in Figure 2. Probability plotting can be a graphical means of determining whether or not the data conform to a distribution based on a subjective visual examination of the data (Hines and Montgomery, 1990). The normal probability plot and histogram of the investigated soil zinc are done by using statistical software SPSS (Norusis, 1996). Figure 3(a) and (b) display this plot and histogram. According to these Figures and Table 1, soil zinc data display a wide variation in space and skewed distribution.

Kriging

Geostatistics (Journel and Hijbregts, 1978; ASCE, 1990) consists of a collection of techniques to analyze spatially correlated data. Techniques such as kriging incorporate the spatial or temporal characteristics of actual data into statistical estimation processes. These techniques can be linear, such as point kriging, ordinary kriging and block kriging.

Geostatistics provides a variogram of data within a statistical framework, including spatial and temporal covariance functions. Unsurprisingly, these models are refereed to as spatial or temporal structures, and are defined in terms

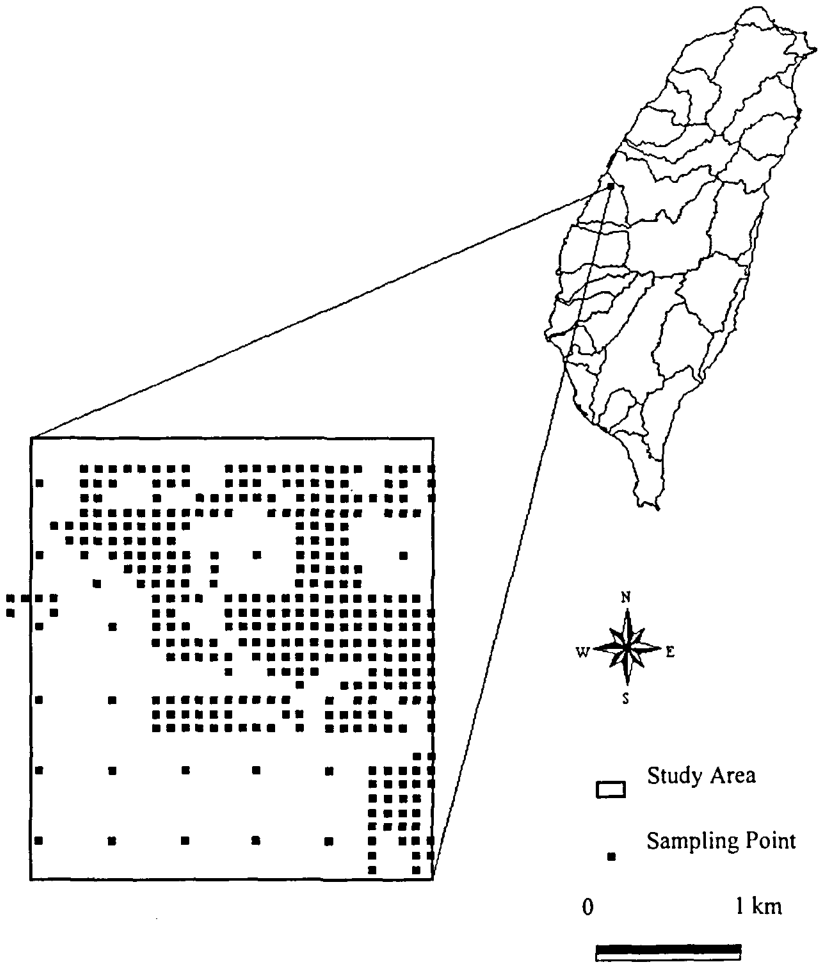


FIGURE 1
Study area and sampling points.

TABLE 1

Descriptive Statistics of Investigated Zn

	Mean	Min	Max	Variance	Median	Kurtosis	Skewness
Zn	80.697 (mg/kg)	11.0 (mg/kg)	1095.0 (mg/kg)	12463.405 $(mg/kg)^2$	47.180 (mg/kg)	29.250	4.582
Ln Zn	3.922 ln(mg/kg)	2.40 ln(mg/kg)	7.00 ln(mg/kg)	0.787 $(\ln(mg/kg))^2$	3.854 ln(mg/kg)	0.116	0.707

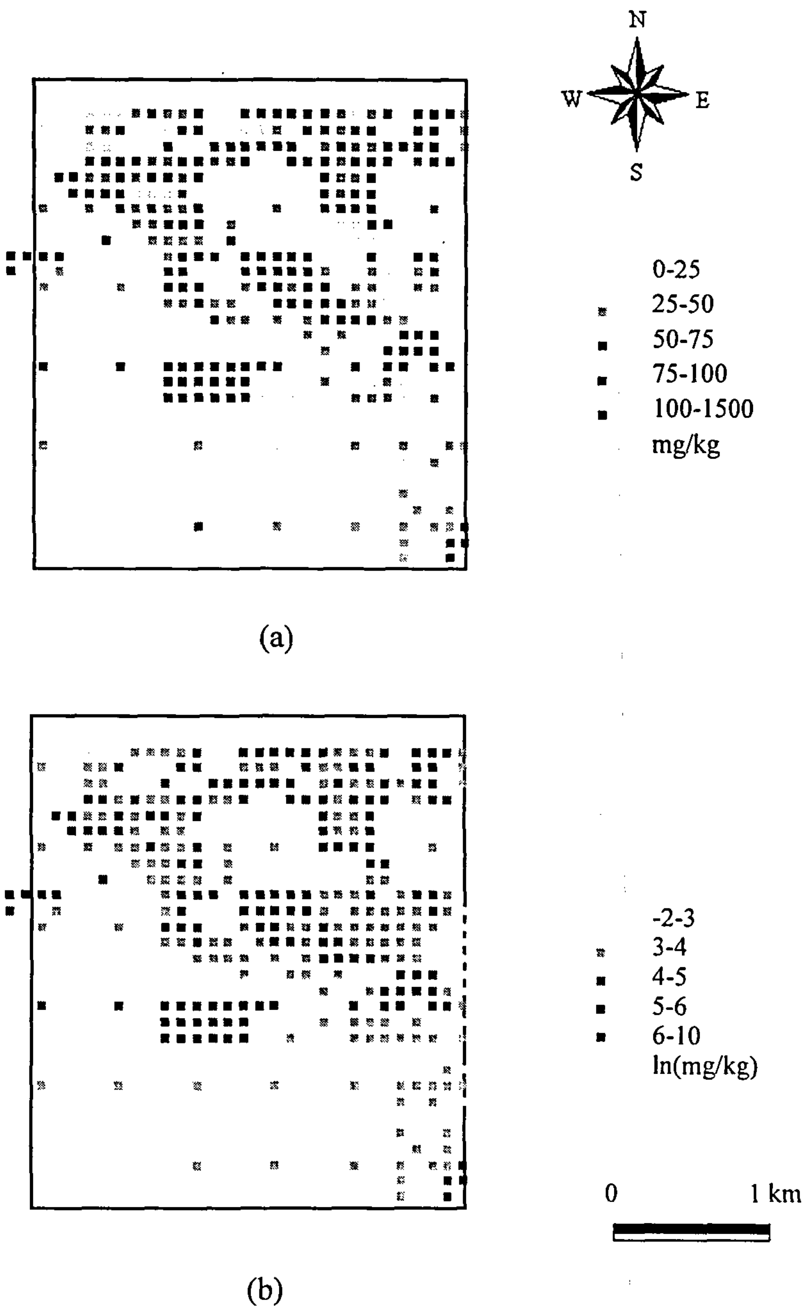
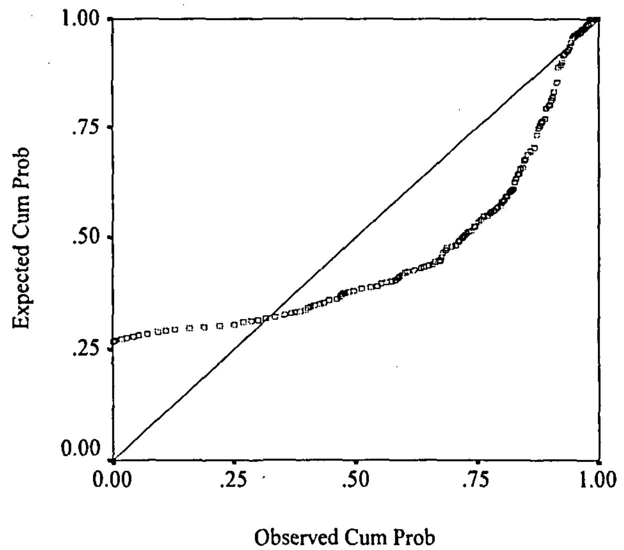
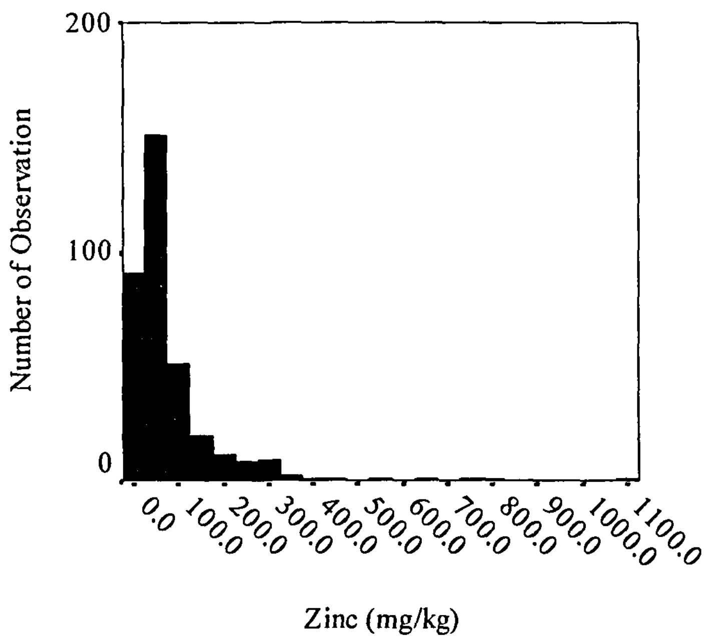


FIGURE 2
Measured values of (a) Zn; (b) ln(Zn).



(a)



(b)

FIGURE 3
(a) Normal probability plot of Zn; (b) histogram of Zn.

of the correlation between any two points separated by either spatial or temporal distances. Kriging estimates are calculated as weighted sums of the adjacent sampled concentrations. These weights depend on the exhibited correlation structure. For illustration, if data appear to be highly continuous in space, those points closer to the estimated points receive higher weights than those farther away. The criterion for selecting these weights is a minimization of the estimation variance. In this framework, kriging estimates can be regarded as the most accurate among all linear estimators (i.e. Best Linear Unbiased Estimator). Therefore, at an unsampled location and given the variogram, a kriging estimate can be thought of simply as an optimally weighted average of the surrounding sampled data (Cressie, 1990).

Sequential Gaussian Simulation

In the Sequential Gaussian Simulation process, simulation is conducted upon the gaussian transformation of the available measurements, such that each simulated value is conditional on the original data and all previously simulated values (Deutsch and Journel, 1992; Rouhani et. al., 1995). A simulated value at visited point is randomly selected from the normal distribution function defined by the kriging mean and variance based on neighborhood values. Finally, the simulated normal values are back transformed into simulated values for an original variable. At the new randomly visited point, the simulated value is conditional on the original data and previously simulated values. This process is repeated until all points are simulated at each realization over the study area.

In this study, NSCORE program in Geostatistical Software Library (GSLIB) (Deutsch and Journel, 1992) is used for fitting a normal variogram of the normalized soil zinc data. The variogram models of Zn, normalized Zn and lnZn are also fit within GS+ (Gamma Design, 1995). SGSIM and OKB2DM in GSLIB perform ordinary kriging, ln kriging and Sequential Gaussian Simulation (SGS) for soil zinc. These simulations and estimations are performed into a square 34 columns by 38 rows grid consisting of 1292, 80m by 80m cells. Five simulations (Sim1, Sim2, Sim3, Sim4, Sim5) are performed at these 1292 cells. The results are also transferred into Arcview 3.0 (ESRI, 1998) for display.

RESULTS AND DISCUSSION

Variograms are calculated for the data at an active lag of 2418m. Least squares model fitting of these variograms generated a relatively consistent set of best-fit models that have the lowest RSS (Model Reduced Sum of Squares) and highest r^2 values. Table 2 lists the parameters from representative models. For investigated values a spherical model with nugget effect=6690(mg/kg)^2, sill=7170(mg/kg)^2 and range=1500m has the best fit among other models (Spherical, or Gaussian) available in the software. The best fit variograms of normalized and natural log investigated values are exponential models with nugget effect=0.484, sill=0.680 and range=694m, and nugget effect=0.38 (ln(mg/kg)^2), sill=7170 (ln(mg/kg)^2) and range=560m as shown in Tables 3 and 4. The variogram models with the high nugget effect illustrate a high small-scale variation or measurement error of investigated data over this study area.

Descriptive Statistics

The ordinary kriging estimates, ln kriging estimates and simulations are based on the above variogram models and 350 observations. Tables 5 and 6 summarize the descriptive statistics of ordinary kriging, ln kriging and Sequential Gaussian Simulation results. The mean value of ordinary kriging estimated values over this study area is extremely close to the mean of investigated values. The median, variance, kurtosis, and skewness of simulations are extremely close to those of investigated data as shown in Table 5. Above results indicate that ordinary kriging process may not preserve the variability of the investigated process. Moreover, this minimization of the prediction error variance involves smoothing the actual variability. Sequential Gaussian Simulations can reproduce the statistics for investigated soil zinc.

To compare the results of ln kriging and Sequential Gaussian Simulation, the simulations are transformed into a natural log format. Similarly, the natural log kriging results display the smoothing effect and low variation on the estimated values as shown in Table 6. SGSs perform multiple realizations that obtain similar statistics to investigated data, as shown in Table 6.

The normal probability plots of simulations and estimations are also done by using statistical software SPSS. Comparing the normal probability plots (Figure

TABLE 2

Variogram Model of Zn

Model	Nugget Effect $(mg/kg)^2$	Sill $(mg/kg)^2$	Range (m)	RSS	r^2
Exponential	5350	8540	482	1.14×10^8	0.351
Spherical	6690	7170	1500	1.03×10^8	0.415
Gaussian	6790	6170	1270	1.04×10^8	0.405

TABLE 3

Variogram Model of Normalized Zn

Model	Nugget Effect (Dimensionless)	Sill (Dimensionless)	Range (m)	RSS	r^2
Exponential	0.484	0.680	694.0	0.05288	0.894
Spherical	0.59	0.54	1833.0	0.05762	0.884
Gaussian	0.66	0.46	1506.0	0.07545	0.848

TABLE 4

Variogram Model of natural log Zn

Model	Nugget Effect $(\ln(mg/g))^2$	Sill $(\ln(mg/kg))^2$	Range (m)	RSS	r^2
Exponential	0.38	0.51	560.0	0.0555	0.849
Spherical	0.48	0.39	1646.0	0.0556	0.848
Gaussian	0.54	0.33	1365.0	0.0672	0.816

TABLE 5

Descriptive Statistics of Kriging estimates and Simulations

	Mean (mg/kg)	Variance $(mg/kg)^2$	Median (mg/kg)	Kurtosis	Skewness
Investigated	80.697	12463.405	47.180	29.250	4.582
Kriging	79.554	2696.932	65.740	3.269	1.750
Sim1	78.108	9883.386	43.273	15.623	3.419
Sim2	82.197	11144.131	48.000	18.695	3.724
Sim3	77.311	10321.810	41.410	15.957	3.499
Sim4	78.491	12762.762	41.000	22.502	4.203
Sim5	74.190	9979.634	41.000	37.602	4.656

TABLE 6

Descriptive Statistics of Ln Kriging Estimates and Simulations

	Mean ln(mg/kg)	Variance $(\ln(mg/kg))^2$	Median ln(mg/kg)	Kurtosis	Skewness
Investigated	3.922	0.787	3.854	0.116	0.707
Kriging	4.202	0.334	4.186	-0.341	0.290
Ln Kriging	3.900	0.253	3.894	-0.246	0.346
Sim1	3.879	0.836	3.767	-0.270	0.634
Sim2	3.941	0.819	3.871	-0.194	0.590
Sim3	3.855	0.858	3.724	0.073	0.587
Sim4	3.846	0.890	3.714	0.761	0.496
Sim5	3.817	0.863	3.714	0.195	0.498

3(a), Figure 4, Figure 5 and Figure 6) reveals that the normal probability plots of the simulated realities more closely fit the normal probability plot of investigated data than that of ordinary kriging estimates.

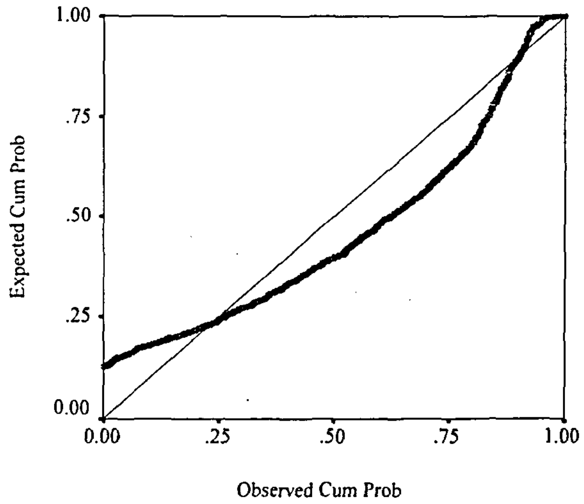
Experimental Variogram

In this study, the experimental variograms of normalized estimated and simulated values are also performed with the same lag interval. According to Figure 7(a), the Sequential Gaussian Simulation can perform very well in terms of reproducing the spatial structure (experimental variogram) for the investigated values. The ordinary kriging values display a well-structured variogram with a low spatial variability and hole effect, but can not perform the spatial structure and small-scale variation for investigated values as shown in Figure 7(a).

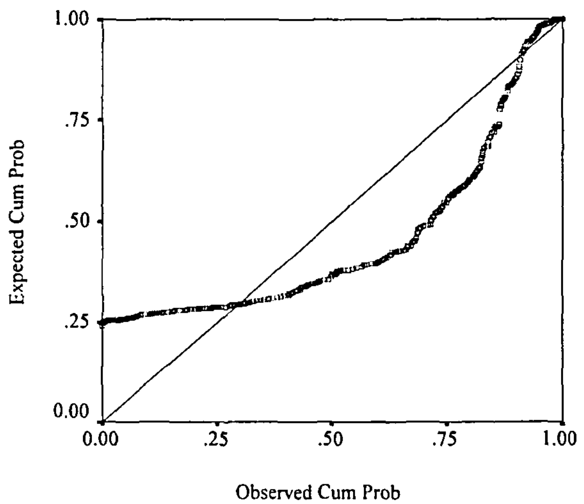
Similarly, the experimental variograms of natural log simulations closely fit the spatial structure for the natural log of investigated values as shown in Figure 7(b). The ordinary kriging estimates of lnZn also display a well-structured variogram with a low spatial variability, but can not perform the small-scale variation for investigated values as shown in Figure 7(b). These results improve the above descriptive statistics results, indicating that simulation can reproduce the statistics of investigated data.

Spatial Distribution

Ordinary kriging and Sequential Gaussian Simulation for Zn in this study area are also performed and mapped in Figure 8. The map of ordinary kriging estimates reveal that kriging tends to smooth out extreme values of the investigated data set. The simulation maps illustrate that the large-scale and small-scale continuity patterns produced by Sequential Gaussian Simulation are visually similar to those in the map of investigated data. Subjectively comparing Figure 2(a) and Figure 8 reveals that the kriging results may overestimate areas with high soil zinc, and underestimate areas with extremely high soil zinc. The conditional simulated maps of Zn are rather irregular, and sites of high concentration in all simulations may have a low concentration at neighboring points as shown in Figure 8. These maps also illustrate that kriging estimates are

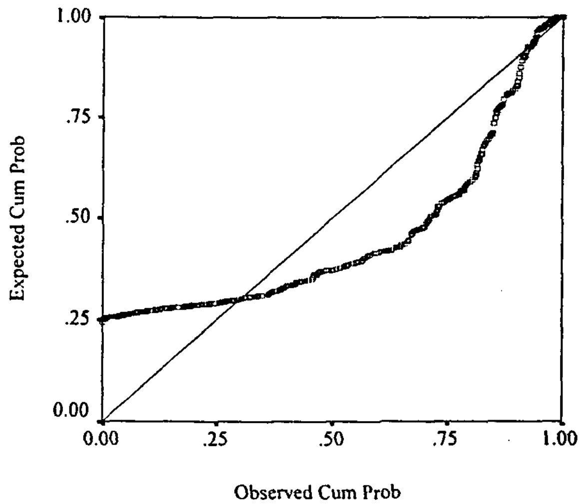


(a)

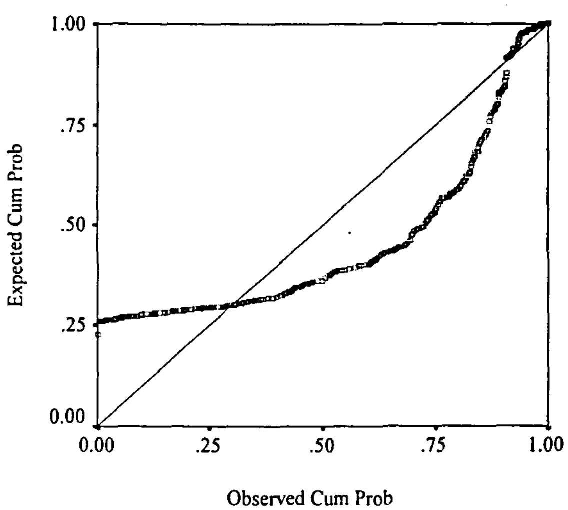


(b)

FIGURE 4
Normal probability plot of (a) ordinary kriging estimation; (b) Sim1.

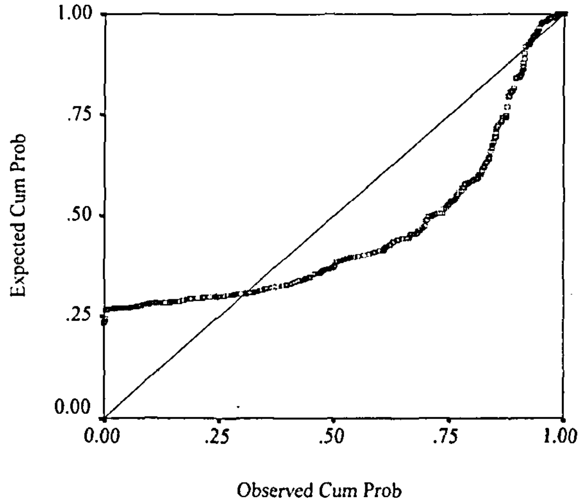


(a)

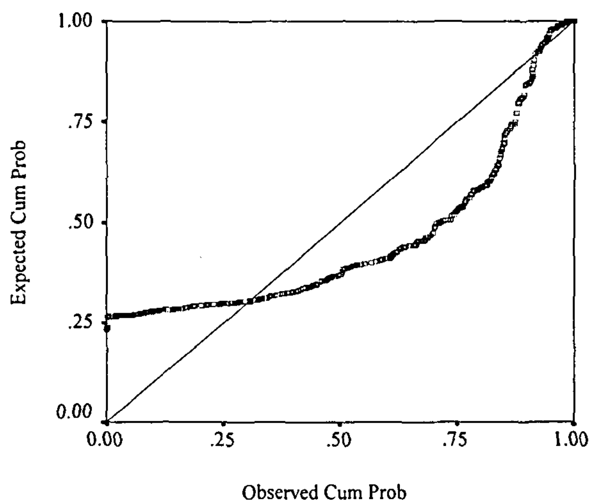


(b)

FIGURE 5
Normal probability plot of (a) Sim2; (b) Sim3.

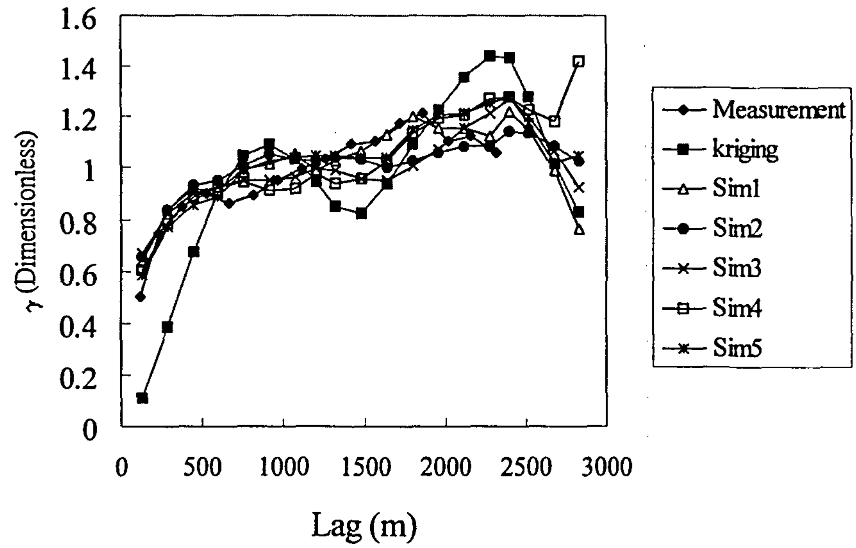


(a)

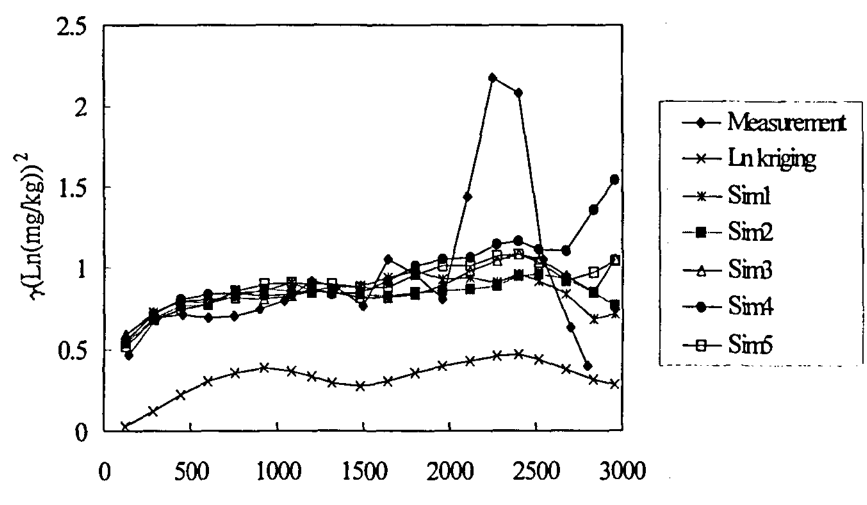


(b)

FIGURE 6
Normal probability plot of (a) Sim4; (b) Sim5.



(a)



(b)

FIGURE 7
Experimental variogram of (a) ordinary kriging estimation and simulations; (b) Ln ordinary kriging estimation and Ln simulations.

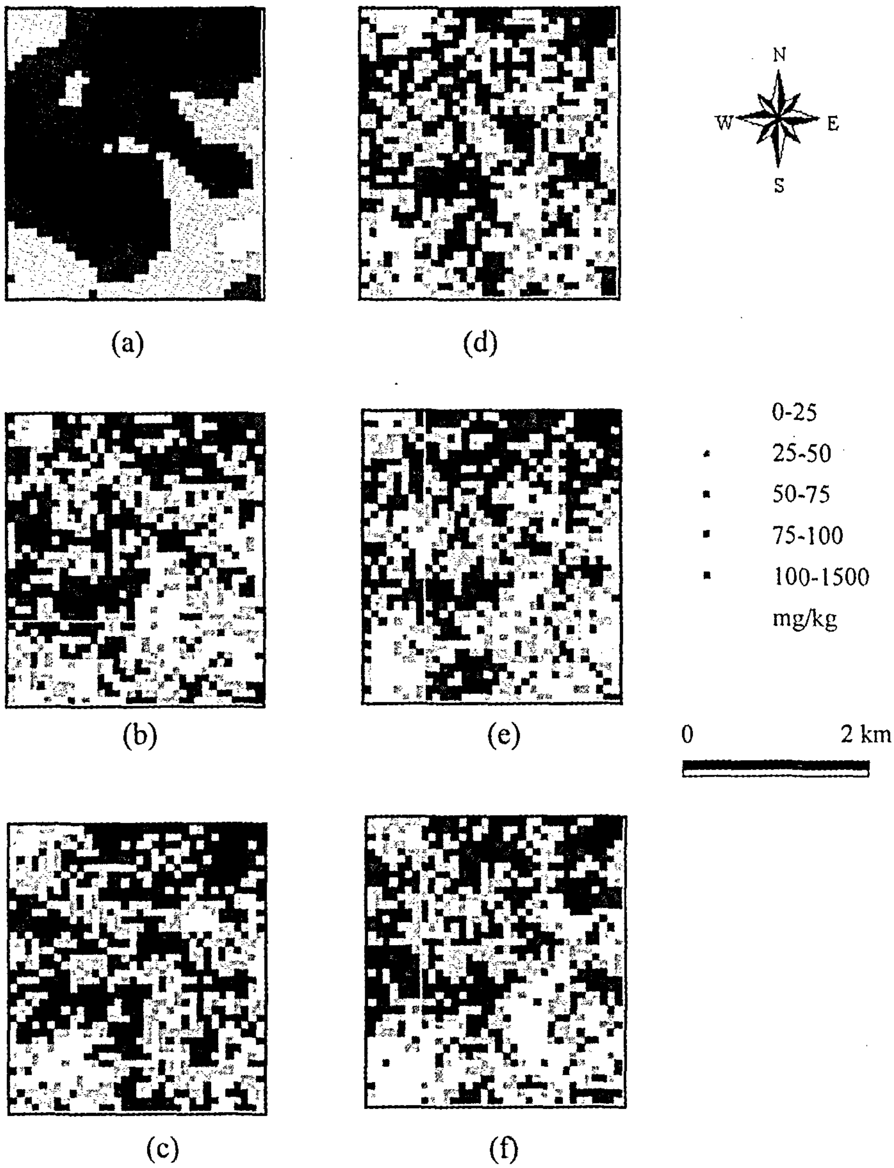


FIGURE 8
Zinc map of (a) ordinary kriging; (b) Sim1; (c) Sim2; (d) Sim3; (e) Sim4; (f) Sim5.

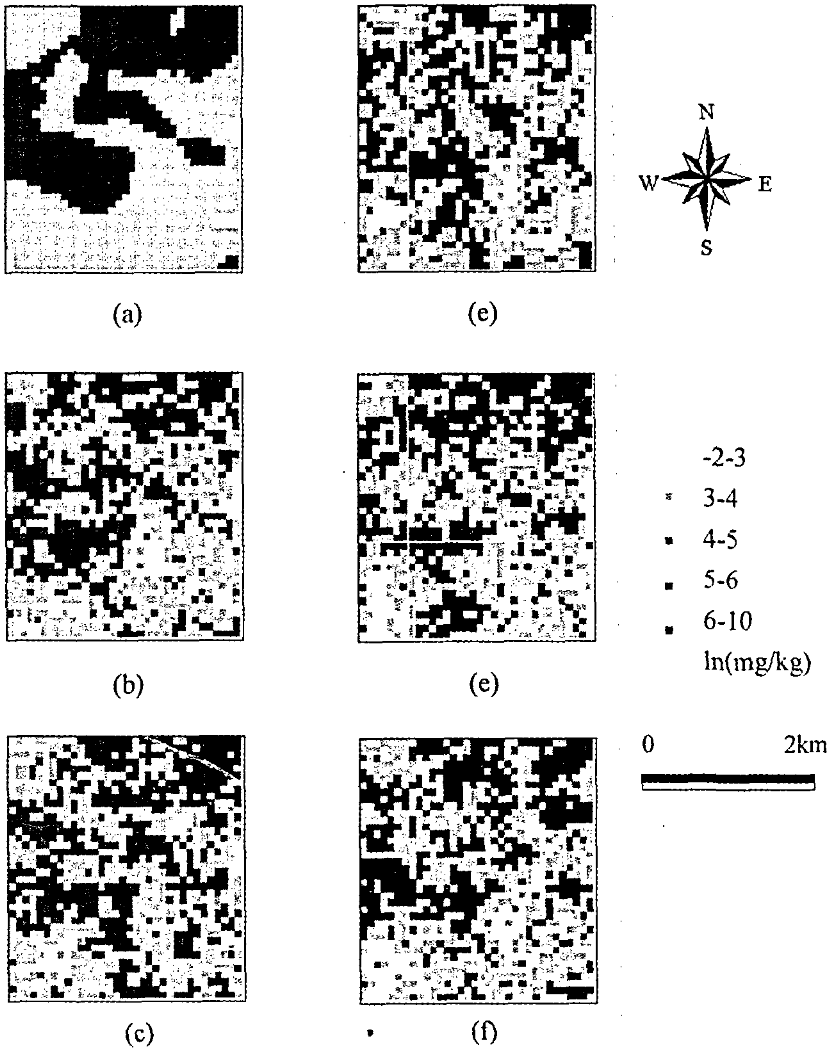


FIGURE 9
Ln zinc map of (a) Ln ordinary kriging; (b) Sim1; (c) Sim2; (d) Sim3; (e) Sim4; (f) Sim5.

much smoother than any of the simulations. For instance, at the central western and northern parts of this study area, kriging map shows a large high concentration area; however, simulation maps do not. However, kriging provides the optimal estimation of Zn at unsampled sites, but do not reproduce the spatial variability for the investigated data in this study case. SGS can reproduce the spatial variation for the investigated data. Moreover, each realization of simulations provides a measure of spatial uncertainty over this study area.

Similarly, ln ordinary kriging results also overestimate areas with high lnZn, and underestimate areas with very high lnZn as shown in Figure 2(b) and Figure 9. The natural log transformed simulations are also mapped in Figure 9. These spatial natural log simulation maps are also rather irregular, and high concentration area in simulations may have low concentration at neighbor sites as shown in Figure 9. These maps also illustrate that natural log kriging estimates are much smoother than simulations.

CONCLUSION

The variogram models with a high nugget effect of investigated soil zinc data illustrate a high small-scale variation or measurement error of investigated data over this study area. Kriging provides the optimal estimation of Zn at unsampled sites. However, the estimated values based on ordinary kriging and natural log kriging display lower variations than the actual investigated soil zinc. These two techniques fail to reproduce measured extremely high soil Zn and ln(Zn) values. Sequential Gaussian Simulation can reproduce both the extreme measured soil zinc and overall zinc spatial distribution. Sequential Gaussian Simulation with multiple realizations has significant advantages at a site with high variation investigated data over ordinary kriging, even natural log kriging techniques. These three techniques may be effective in assessing the uncertainty of investigated data. Spatially, Sequential Gaussian Simulation can also be used to assess stochastic elements in a complex soil heavy metal study.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 87-2621-P-002-012.

REFERENCES

ASCE Task Committee on Geostatistical Techniques in Geohydrology (S. Rouhani, Chairman and Principle Author), 116(5), 612-632 (1990).

Couto E. G., Stein, A. and Klamt, E., Agriculture, Ecosystems and Environment, 66, 139-152 (1997).

Cressie C,, Math. Geology, 22, 239-252 (1990).

Deutsh C. and Journel, A. G., GSLIB, Geostatistical Software Library and User's Guide, Oxford University Press., New York (1992).

Dowd, P. A., Math. Geology, 14, 474-500 (1982).

Environmental Systems Research Institute, Arcview GIS, ESRI, Redlands, CA (1996).

Gamma Design Software, GS+: Geostatistics for the agronomic and biological sciences. Version 2.3. Gamma Design, Plainwell, MI, (1994).

Gotway, C.A. and Hergert, G.W., Soil Sci. Am. J., 61, 298-309 (1997).

Gotway C. A., Ferguson, R. B., Hergert, G. W., and and Peterson, T. A., Soil Sci. Am. J., 60, 1237-1247 (1996).

Hines, W. W. and Montgomery, D. C., Probability and Statistics in Engineering and Management Science, (1990).

Isaaks, E. H. and Journel, A. G., Applied Geostatistics, Oxford University Press Inc. (1992).

Juang, K. W. and Lee, D. Y., J. Environ. Qaul., 27, 355-363 (1998).

Journel, A. G. and Huijbregts,C. J., Mining Geostatistics, Academic Press. , (1978).

Journel, A. G., Math. Geology, 12, 285-301 (1980).

Keck, T.J., Quimby, W.F. and Nielson, G.A., Soil Sci. Soc. Am. J., 57, 782-786 (1993).

Kentwell, D. J., Bloom, L. M. and Comber, G. A., Math. Geology, 31, 311-325 (1999).

Litaor, M.I., J. Environ. Qual., 24, 506-516 (1995).

Mowrer, H. T., Ecological Modelling, 98, 73-86 (1997).

Norusis, M. J., SPSS for Windows Base System User's Giude Release 6.0, (1993).

Rendu, J.-M., Math. Geology, 11, 407-422 (1979).

Rivoirard, J., Math. Geology, 22, 213-221 (1990).

Roth, C., Math. Geology, 30, 999-1009 (1998).

Rouhani, S., Lin, Y. P. and Shi, Y., H-Area/ITP Geostatistical Assessment In-situ and Engineering Properties, Final Technical Report, Georgia Institute of Technology (1995).

Samra, J.S. and Gill, H.S., Soil Sci., 155, 148-153 (1993).

Sterk, G. and Stein, A., Soil Sci. Soc. Am. J., 61, 232-239 (1997).

Wang, X. J., J. of Environ. Sci. Health, Part A, 33, 1057-1073 (1998).

Wang, X. J. and Zhang, Z. P., J. of Environ. Sci. Health, Part A, 34, 73-89 (1999).

White J. G., Welch, R. M. and Norvell,W. A., Soil Sci. Soc. Am. J., 61, 185-194 (1997).

Yost, R.S., Uehara, G. and Fox,R. L., Soil Sci. Am. J., 46, 1033-1037 (1982).

Received: July 16, 1999